(NASA-TM-101275) WORKLOAD CONSULTANT: A MICROPROCESSOR-BASED SYSTEM FOR SELECTING WORKLOAD ASSESSMENT PROCEDURES (NASA) 6 p

N90-70379

Unclas
00/60 0252394

WORKLOAD CONSULTANT : A MICROPROCESSOR-BASED SYSTEM FOR SELECTING WORKLOAD ASSESSMENT PROCEDURES

Patricia A. Casper
Purdue University
West Lafayette, IN

Robert J. Shively
U. S. Army
Aeroflightdynamics Directorate
NASA - Ames Research Center
Moffett Field, CA

Sandra G. Hart
NASA - Ames Research Center
Moffett Field, CA

Abstract

Recent years have seen a deepening interest in the measurement of human operator workload. However, not all persons involved in the design and production of human-machine systems are educated in the rigors of workload measurement and the currently available techniques. Furthermore, as in most areas of expertise, there aren't enough human "experts" to go around. The present paper describes an "expert" system, created at the NASA-Ames Research Center, that was designed to provide decision support for persons interested in assessing operator workload. The system is based on current research in the field of workload measurement and is flexible enough to allow for incorporation of new knowledge as it is empirically validated.

WC FIELDS (Workload Consultant for Field and Industrial Environments for Laypersons Desiring Support) is a microprocessor-based system designed to aid users in selecting appropriate workload assessment procedures for their particular application. The system is designed to aid non-experts in the area of workload. These users may be new to the field, or interested in workload only as a control variable and not as a primary focus of their research. Presently, for these researchers to apply workload methodology they must either consult with an expert or sort through numerous books and articles on workload assessment and theory. Each of these sources tends to focus on the favorite assessment technique of the author or expert, and is usually applied to specific situations. Consequently, sorting through the literature to find measures appropriate to one's application becomes a very difficult task for those not familiar with the field (and often for those who are familiar with the field).

WC FIELDS is designed to suggest appropriate workload assessment techniques, taking into account a wide variety of workload questions, environments, and assessment methodologies. The system is intended to take the place of a workload expert in suggesting workload assessment methodologies and to provide enough information and appropriate references for researchers to apply the suggested methods. The system will not suggest a single assessment procedure as the one of choice; rather it will present the user with several assessment methodologies appropriate for his/her application and sufficient information for the user to make an intelligent choice among alternatives.

Structure of the "expert" system

The system was developed using EXSYS, a commercially-available rule-based expert system development package. The knowledge base was created by compiling an extensive list of IF - THEN type rules which the expert system translates into questions to ask the user. The rules are structured in the form:

IF condition X is Y THEN

1. method a : probability = 9/10
2. method b : probability = 7/10
3. method c : probability = 3/10

where the probability ranges from 0 (absolutely not appropriate) to 10 (absolutely appropriate) and refers to the appropriateness of the given workload assessment method for the given application. The user of the system does not see the rule but sees instead a question derived from the rule by the program. The question formed by the program for the above rule would look as follows:

Condition X is

1. W
2. X
3. Y
4. Z

Among the more desirable features of this system is it's ability to use backward chaining (the program can obtain information from the rules in the knowledge base rather than continually prompting the user for information previously obtained). The following paragraphs describe how WC FIELDS operates:

Initially, the user is presented general information about the program. The user then may choose whether or not he/she wishes to have the rules displayed as the program verifies that they are true. Knowledge of the rules used to reach conclusions can help a user understand why certain workload assessment methodologies are selected, and in some instances, why the rule can be interpreted in different ways depending on a user's special applications. Occasionally, brief notes to the user are presented to justify the rationale underlying a particular rule.

After the introductory questions, the user is asked a number of multiple choice questions pertaining to the goals of the project for which workload will be measured and to the environment in which workload will be measured. For each question the user has the option of selecting one or more of the available answers. If the user wants to know why a question is being asked, he/she need only respond "WHY" instead of choosing one of the alternative choices, and the rule that prompted the question will be displayed. This option, available even if the user initially chose not to have the rules displayed, may be exercised at any point in the program.

When the program has obtained enough information to reach a decision, the possible outcomes are numerically evaluated using the probabilities associated with choices in the rules. User responses with a value of 0 or 10 permanently "lock" the value of the choice, while values of 1 to 9 are averaged over all of the rules that used the choice.

Before the final results are presented, a screen of information advising the user on the meaning of the numerical values assigned to the various measures appears on the screen. The next screen presents each of the possible workload measures in order of their appropriateness, with only the numerators of the probability ratios given. Choices to which no rule applied or for which a value of 0 was obtained will not automatically be displayed, but may be displayed if the user so indicates. The user may then exercise his/her own best judgment concerning acceptable probability levels necessary to choose a measure or measures.

When the final choices appear, the user may change one or more of the answers previously given to any of the questions and test the effect that those changes have on the final outcome. This option is especially useful for comparing outcomes when questions are answered differently and may be done any number of times. The answers given may also be saved on disk to be used in a subsequent run of the program at a later time. The results of a run may also be printed with or without the information used in reaching the final results.

Also available to the user is the option of obtaining additional information on any of the measures displayed on a menu. This "information base" covers several of the most thoroughly validated and commonly used examples of each type of workload measure. Included is a description of each technique and the environments in which it is commonly used, instructions about how to implement it, and a brief review of it's strengths and weaknesses. The description of each measure is accompanied by a list of references so that the user can find further information should he/she decide to implement or become more familiar with the measure. The next section of this paper briefly outlines the content of the workload knowledge base of WC FIELDS.

Knowledge base content

For each combination of workload question and environment, a set of the most appropriate measures will be suggested. The measures were selected on the basis of their validity, reliability, and frequency of use. As new measures become available, they may be added to the system once they have been validated. The measures are divided into five categories: primary task performance, secondary task performance, subjective ratings, physiological measures, and analytic techniques. The following paragraphs review the measures that are presently included in the system.

Primary task performance

This class of measures assumes that an operator's performance on the given task is by itself a valid indicator of workload. For example, if a pilot can successfully land the plane he is flying then the workload for the landing task must have been within acceptable levels. Primary task measures are often difficult to identify, as they are often task- or situation-dependent. Several measures have been found to be useful indicators of workload, however, such as frequency and content of communications, manual control variability, errors, and the ability to meet schedules.

Primary task performance measures have the advantage of face validity, that is, performance on the primary task appears to accurately reflect the operator's ability to perform the task. Occasionally designers accept primary task performance measures as indicators of task workload and don't attempt to apply other methods. While this is a necessary measure of total task performance, it often is not sufficient to measure workload. The operator may be overloaded while performing a task, but, through increased effort, can maintain criterion performance for some time until performance rapidly declines -- a dangerous situation in some environments. In addition, a primary task measure gives no indication of the ability of the operator to perform additional tasks as are often necessary in emergency or peak workload situations. Finally, performance measures rarely reflect workload variations at the low end of the workload range. Performance decrements usually occur only in response to relatively high levels of workload.

Secondary task performance

Secondary task measures of workload are obtained by giving the operator an "extra" task to perform in addition to his primary task. The operator is instructed to maintain a specified level of performance on the primary task (e.g. flying a plane) and to use any remaining capacity (effort, resources, time, etc.) to perform the secondary task to the best of his ability. The level of performance on the secondary task is assumed to indicate the amount of "spare" capacity not demanded by the primary task. The secondary tasks included in the system are: time estimation (Gunning, 1978; Hart, McPherson, and Loomis, 1978; Bortolussi, Kantowitz, and Hart, 1985), the Sternberg memory task (Sternberg, 1969, 1975), manual tracking (Brickner and Gopher, 1981; Wickens, Sandry, and Vidulich, 1983), auditory and visual choice reaction time (Kantowitz, Hart, and Bortolussi, 1983; Kantowitz, Hart, Bortolussi, Shively, and Kantowitz, 1984; Kantowitz and Knight, 1976a), monitoring (Ephrath and Curry, 1977), mathematical calculations, and dichotic listening (Gundry, 1976; McLeod, 1973).

These measures have the advantage of being highly diagnostic in an information processing paradigm. In other words, they have the ability to identify the cognitive processes that are affected by the demands of the imposed tasks. Secondary task methodology is backed by a strong theoretical base from cognitive psychology. The problem is that no secondary task measure is universally applicable, because different types of processing resources may be consumed (or remain) depending on the nature of the demands imposed by a primary task. Performance on one secondary task may indicate that no additional resources remain to perform an additional activity (indicating high workload), while another would show no such limitations (indicating low workload) because they do, or do not, compete for the same limited resources. Thus, the user must select a secondary task that will tap the same resources as the task of interest. In doing so, much of the ability to compare results across experiments is lost.

In addition, implementation of secondary tasks in an operational environment may be problematic in that performance of the additional task may interfere with performance of the primary task. For this reason, researchers have attempted to use embedded secondary tasks. These tasks are designed to have the appearance of being an integrated part of the primary task. An example of an embedded secondary task would be increasing communication task loadings in an aviation environment. The communications represent a task that pilots must normally perform and thus seem realistic, however they also represent a secondary task in relation to the primary task of flying the aircraft.

Subjective rating scales

Subjective ratings are usually obtained by asking the operator either during or after a task how mentally taxing the task was to perform. Ratings can come from verbal protocols or from paper and pencil questionnaires, and can be obtained either from the operator him/herself or from an external observer who estimates the workload of the operator performing the task. Subjective ratings are intuitively appealing in that there should be no easier way to find out how hard someone is working on a task than to ask him.

Several types of subjective rating scales are included in the system. They range in complexity from a unidimensional numeric scale to a multidimensional weighted average of several component factors. Ratings may focus solely on task-related mental workload or may also encompass psychological factors such as level of physical or physiological arousal. Several rating techniques, like SWAT (Subjective Workload Assessment Technique), rely on subjective estimates of the importance of component dimensions to the subject's a priori definitions of workload (Eggemeier, 1981; Eggemeier, Crabtree, Zingg, Reid, and Shingledecker, 1982; Reid, Shingledecker, Nygren, and Eggemeier, 1981; Reid, Eggemeier, and Nygren, 1982). The NASA bipolar rating scale (Hart, Battiste, and Lester, 1984) incorporates the operators' ratings of the relevance of each factor in a task (e.g., frustration, physical demand, mental demand, performance, effort, and temporal demand) into a weight associated with the rating for each factor. The final workload score is a weighted average of the ratings from each scale. Other multidimensional techniques rely on a hierarchy of questions about different task-related factors that lead the rater to a single numeric workload rating (e.g., the Cooper-Harper Handling Qualities Rating Scales, and the modified Cooper-Harper Scale [Cooper and Harper, 1969; Sheridan and Simpson, 1979; Wierwille, Skipper, and Rieger, 1984]).

Also included in the system are techniques relying on reference tasks (Gopher and Braune, 1984) or reference vehicles (Ruggiero and Fadden, 1982), and techniques based on linguistic descriptors of component factors that are combined with fuzzy set theory (Moray, 1983).

Of all the measures currently available this class is no doubt the easiest to implement, however, it does have several drawbacks. Theoretical support of ratings techniques is not as strong as with some other measures (secondary tasks, for example), ratings obtained by experimenters using different scales are difficult to compare, and scaling problems result from the use of relative rather than absolute numerical data. Nonetheless the popularity of this technique continues and overall the advantages outweigh the disadvantages.

Physiological measures

The physiological indices of mental workload are the most objective measures in terms of their ability to indicate operator's reactions to demands placed on them. Two classes of physiological measures are included in the system: those that measure the overall level of activation or arousal and those that relate more directly to mental and perceptual processing. Included in the first group are heart rate (Roscoe, 1982; Hart and Hauser, 1984), heart rate variability (Casali and Wierwille, 1982, 1983; Wierwille and Williges, 1978), galvanic skin response (O'Donnell, 1979), and respiration. Measures relating more directly to cognitive processes are evoked cortical potentials (Israel, Wickens, Chesney, and Donchin, 1980), eye blinks (Stern and Skelly, 1984), spectral analysis of heart beat to beat intervals (Mulder, 1979), eye point of regard and fixation duration (Sanders, Hofmann, Simmons, and DeBonis, 1979), and pupil diameter (Casali and Wierwille, 1983).

Physiological measures are advantageous in that they are sensitive to high levels of workload and are unlikely to interfere with performance of the primary task. Unobtrusiveness is a very desirable feature in some operational environments where disruption of performance on a primary task (such as flying a helicopter in a nap of the earth, or NOE, mission, for instance) could mean serious injury or death to the operator for whom workload is being measured.

Drawbacks related to the use of these measures include the enormous magnitude of data created by the continuous monitoring of biological functions and the noise embedded in the data from such functions. Computing facilities to support such analyses are expensive and time consuming to maintain and operate. A further problem is the current lack of a sound theoretical framework linking physiological processes to their cognitive correlates.

Analytic techniques

There are several computer programs that compute workload using task and time-line analytic techniques and human operator models. Most have been developed for specific applications, such as the design of a new aircraft, and thus do not generalize to other applications (McCraken and Aldrich, 1981). Furthermore, most are not in the public domain (as are all of the other measures described), and are generally available only through purchase or contract. The techniques included in WC FIELDS are time-line analysis and task analysis.

When using techniques based upon task and timeline analysis, the researcher identifies the tasks and the times to perform the tasks either through expert opinion or empirical techniques. The tasks are then placed on a timeline and the measure of workload is the ratio of time required for task performance to the time available for task performance. A drawback to this technique is that in practice, operators will reschedule tasks to reduce workload peaks and thus the predicted workload peaks may not occur. In addition, tasks are rarely strictly additive in terms of time to perform. In some cases operators integrate tasks in order to reduce the total time necessary to perform both tasks. In other cases the tasks may interfere with each other and the total time necessary to perform both tasks would be more than that predicted by a strictly additive model.

Conclusions

Careful consideration of human operator workload prior to implementation of newly designed man-machine systems has become a primary objective in many areas of industry. Workload assessment is vital to industries supporting man/machine interaction for economic, safety, and productivity reasons. It is hoped that the "expert" system described in this paper will be a valuable assistant involved in system design from the initial conceptualization to the final validation stages.

References

Bortolussi, M. R., Kantowitz, B. H., and Hart, S. G. (1985) Measuring pilot workload in a motion base trainer : A comparison of four techniques. Proceedings of the Third Biannual Symposium on Aviation Psychology. Columbus, OH : Ohio State University, 263-270.

Brickner, M., and Gopher, D. (1981) Improving time-sharing performance by enhancing voluntary control on processing resources. (AFOSR-77-3131C) Washington, D. C. : Air Force Office of Scientific Research.

Casali, J. G., and Wierwille, W. W. (1982) A sensitivity/intrusion comparison of mental workload estimation techniques using a flight task emphasizing perceptual piloting activity. Proceedings of the 1982 IEEE International Conference on Cybernetics and Society. New York : Institute of Electrical and Electronic Engineers, 598-602.

Casali, J. G., and Wierwille, W. W. (1983) A comparative evaluation of rating scale, secondary task, physiological, and primary task workload estimation techniques in a simulated flight task emphasizing communications load. Human Factors, 25(6), 623-641.

Cooper, G. E., and Harper, R. P. (1969) The use of Pilot Rating in the Evaluation of Aircraft Handling Qualities (NASA TN-D-5153) Washington, D. C. : National Aeronautics and Space Administration.

Eggemeier, F. T. (1981) Current issues in subjective assessment of workload. Proceedings of Human Factors Society 25th Annual Meeting. Santa Monica, CA : Human Factors Society, 513-51?.

Eggemeier, F. T., Crabtree, M. S., Zingg, J. J., Reid, G. B., and Shingledecker, C. A. (1982) Subjective workload assessment in a memory update task. Proceedings of the Human Factors Society 26th Annual Meeting. Santa Monica, CA : Human Factors Society, 643-647.

Ephrath, A. R., and Curry, R. E. (1977) Detection by pilots of system failures during instrument landings. IEEE Transactions on Systems, Man, and Cybernetics, SMC-7(12),841-848.

Gopher, D., and Braune, R. (1984) On the psychophysics of workload : Why bother with subjective measures? Human Factors, 26(5), 519-532.

Gundry, A. J. (1976) Thresholds to roll motion in a flight simulator. Proceedings of the Visual and Motion Simulation Conference, Dayton, Ohio, American Institute of Aeronautics and Astronautics, April, 1-9.

Gunning, D. (1978) Time estimation as a technique to measure workload. Proceedings of the Human Factors Society 22nd Annual Meeting. Santa Monica, CA : Human Factors Society, 41-45.

Hart, S. G., Battiste, V., and Lester, P. T. (1984) POPCORN : A supervisory control simulation for workload and performance research. Proceedings of the 20th Annual Conference on Manual Control. (NASA CP 2341) Washington, D. C. : National Aeronautics and Space Administration, 431-454.

Hart, S. G., and Hauser, J. R. (in press) Inflight application of three pilot workload measurement techniques. Aerospace and Environmental Medicine.

Hart, S. G., McPherson, D., and Loomis, L.L. (1978) Time estimation as a secondary task to measure workload : Summary of research. Proceedings of the 14th Annual Conference on Manual Control. Los Angeles, CA : University of Southern California, 693-712.

Israel, J. B., Wickens, C. D., Chesney, G. L., and Donchin, E. (1980) The event-related brain potential as an index of display-monitoring workload. Human Factors, 22(2), 211-224.

Kantowitz, B. H., Hart, S. G., and Bortolussi, M. (1983) Measuring pilot workload in a motion-based simulator : I. Asynchronous secondary choice-reaction task. Proceedings of the Human Factors Society 26th Annual Meeting. Santa Monica, CA : Human Factors Society.

Kantowitz, B. H., Hart, S. G., Bortolussi, M. R., Shively, R. J., and Kantowitz, S. C. (1984) Measuring pilot workload in a moving-base simulator : II. Building levels of workload. Proceedings of the 20th Annual Conference on Manual Control. (NASA CP2341) Washington, D. C. : National Aeronautics and Space Administration, 359-372.

Kantowitz, B. H., and Knight, J. L. (1976a) Testing tapping timesharing, II : Auditory secondary task. Acta Psychologica, 40, 343-362.

McCraken, J. H., and Aldrich, T. B. (1981) Analyses of selected LHX mission functions : Implications for operator workload and system automation goals. Anacapa Sciences, Inc.

McLeod, P. D. (1973) Interference of "attend to and learn" tasks with tracking. Journal of Experimental Psychology, 99, 330-333.

Moray, N. (1983) A fuzzy model of rather heavy workload. Proceedings of the 18th Annual Conference on Manual Control (AFWAL-TP-83-3021) Wright-Patterson Air Force Base, OH : Air Force Wright Aeronautical Laboratories.

Mulder, G. (1979) Sinusarrythmia and mental workload. In Mental Workload : Its Theory and Measurement. N. Moray (Ed.) New York : Plenum Press, 327-344.

O'Donnell, R. D. (1979) Contributions of Psychophysiological Techniques to Aircraft Design and other Operational Problems. (AGARDOGRAPH 244). London : Harford House.

Reid, G. B., Shingledecker, C. A., Nygren, T. E., and Eggemeier, F. T. (1981) Development of multidimensional subjective measures of workload. Proceedings of the International Conference on Cybernetics and Society, New York : Institute of Electrical and Electronic Engineers, 403-406.

Reid, G. B., Eggemeier, F. T., and Nygren, T. E. (1982) An individual differences approach to SWAT scale development. Proceedings of the Human Factors Society 26th Annual Meeting. Santa Monica, CA : Human Factors Society, 639-642.

Roscoe, A. H. (1982) Heart rate as an in-flight measure of pilot workload. Proceedings of the Workshop on flight Testing to Identify Pilot Workload and Pilot Dynamics. (AFFTC-TR-82-5) Edwards AFB, CA, 338-349.

Ruggiero, F. T., and Fadden, D. M. (1982) Pilot subjective evaluation : A practical supplement to traditional assessments of workload. Proceedings of the International Conference on Cybernetics and Society. New York : Institute of Electrical and Electronics Engineers, 603-607.

Sanders, M. G., Hofmann, M. A., Simmons, R. R., and DeBonis, J. N. (1979) Visual workload of the copilot/navigator during terrain flight. In Studies on Pilot Workload. R. Auffret (Ed.) (AGARD Conference Proceedings No 217) London : Harford House.

Sheridan, T. B., and Simpson, R. W. (1979) Toward the Definition and Measurement of the Mental Workload of Transport Pilots. (Final Report DOT-OS-70055) Cambridge : Massachusetts Institute of Technology.

Stern, J. A., and Skelly, J. J. (1984) The eyeblink and workload considerations. Proceedings of the Human Factors Society 28th Annual Meeting. Santa Monica, CA : Human Factors Society, 942-944.

Sternberg, S. (1969) The discovery of processing stages : Extensions of Donders' method. Acta Psychologica, 30, 276-315.

Sternberg, S. (1975) Memory scanning : New findings and current controversies. Quarterly Journal of Experimental Psychology, 27, 1-32.

Wickens, C. D., Sandry, D. L., and Vidulich, M. D. (1983) Compatibility and resource competetion between modaiitites of input, central processing, and output. Human Factors, 25(2), 227-248.

Wierwille, W. W., Skipper, J. H., and Rieger, C. A. (1984) Decision tree rating scales for workload estimation : Theme and variations. Proceedings of the 20th Annual conference on Manual Control. (NASA CP 2341) Washington, D. C. : National Aeronautics and Space Administration, 73-84.

Wierwille, W. W., and Williges, R. C. (1978) Survey and Analysis of Operator Workload Assessment Techniques. (Report S-78-101) Blacksburg, VA : Systemetrics.